

# Data curation implementation at the Human Sciences Research Council

## Case study

Dr Lucia Lötter

# Human Sciences Research Council

## Mission

The HSRC is a statutory research organisation that advances social sciences and humanities for public use



# What we do

## ± 200 research projects annually

- Reduce poverty, grow economy and create jobs
- Improve quality of education
- Promote human and social advancement through skills development
- Accelerate service delivery
- Reduce crime
- Develop youth, families and social cohesion
- Promote health, health systems and wellbeing
- Investigate HIV/AIDS and STI control
- Provide advice on nutrition and food security
- Interpret and understand our national system of innovation





# HSRC Research Data Archive

## Vision

Our vision is to be a **digital repository facility for the HSRC's research data** in support of evidence based human and social development in South Africa and the broader region.

## Mission

To make research data accessible and to ensure its future survival and usability by managing the data sets as an effective information source, and

- ensuring the long time **preservation** of these digital resources, and
- **disseminating** the data for secondary use.

to inform monitoring and evaluation, further analysis, as well as debate, advocacy decision-making and training.

# HSRC Research Data Archive

## Objectives

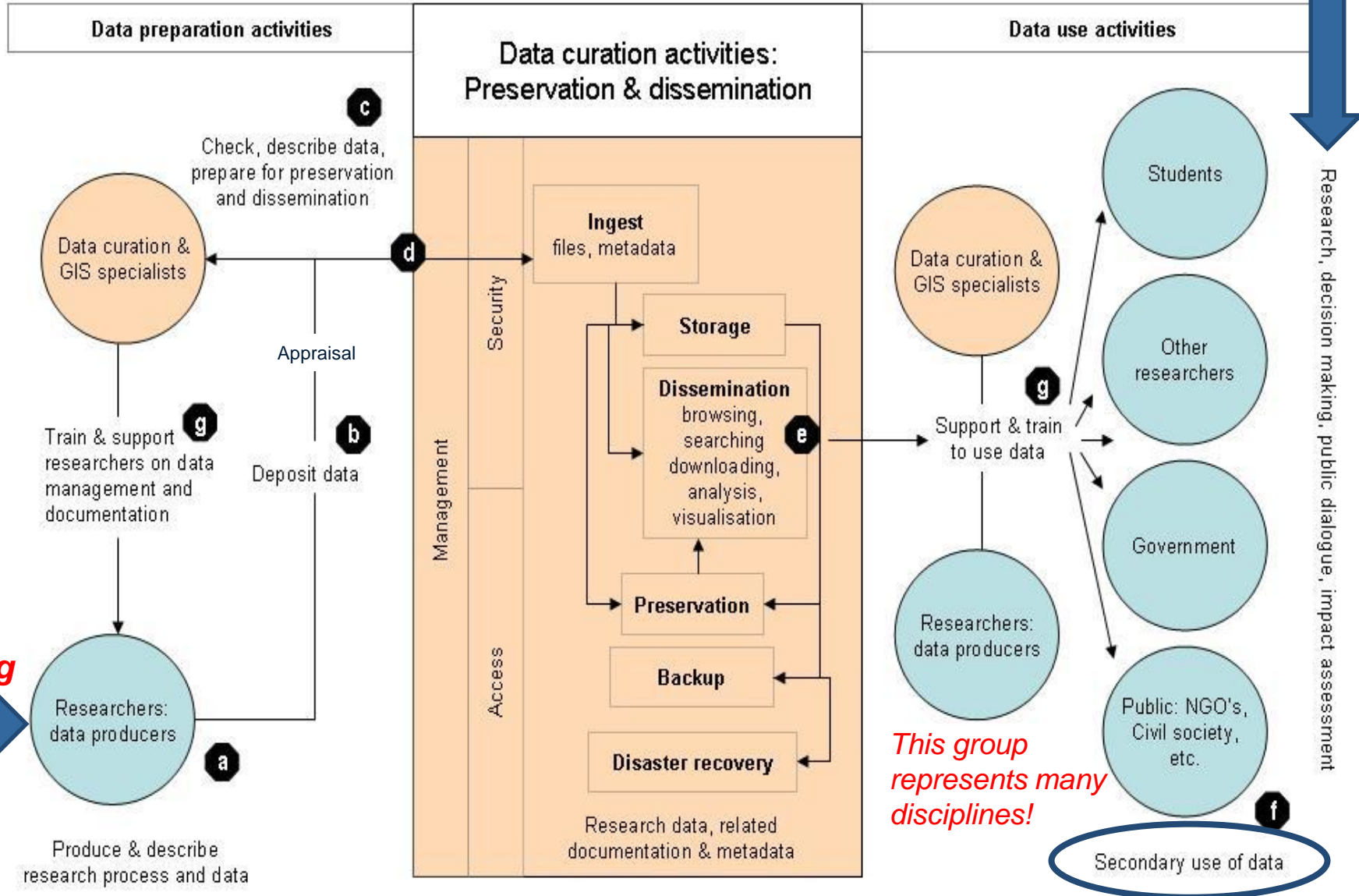
- Establish a repository for quantitative and qualitative data in the Social Sciences in SA
- Develop a data collection of high re-use value that speaks to the priorities of a developing country
- Promote best practice in terms of data management standards
- Promote the use and impact of the data collection in the archive
- Build meaningful and enduring relationships with stakeholders – nationally and internationally
- Demonstrate excellence in accountability and operations

# Nature of the HSRC's data collection

- Spans a variety of content areas in the Social Sciences and Humanities
- Quantitative and qualitative data
  - Focus on cross-sectional repeat surveys (at least three successive “waves” for meaningful trends)
  - Many mixed method studies
- Owned by the HSRC or external organisations

*With the end in mind ...*

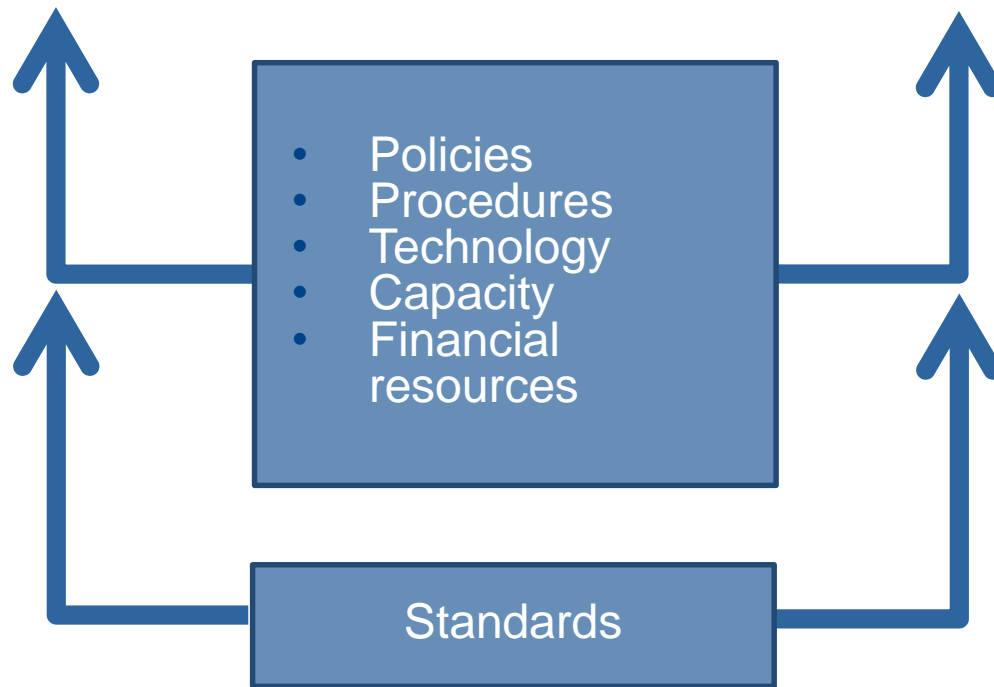
# Data curation - A means towards an end





# What should be in place?

- Collection development
- Depositor support
- Digital object management
- Promote and facilitate secondary data discovery and use



# What should be in place?

- Collection development
- Depositor support
- Digital object management
- Promote and facilitate secondary data discovery and use

Organisational context

Policies  
• Procedures  
• Technology  
• Capacity  
• Financial resources

Standards

*No initiatives will (or should) be exactly the same!!!*

# What is the context of the curation process?

- **Bottom-up**
  - What is research and data all about?
  - What are researcher expectations?
  - How can researchers be supported?
  - How do researchers manage and document data?
  - What are the domain specific ontology, thesaurus, or metadata scheme?
  - Which data repository services do the researchers use?
- **Top-down**
  - What is the organisations' priorities?
  - What is the aim of the repository?
  - What is organisational commitment, policies, procedures, resources?

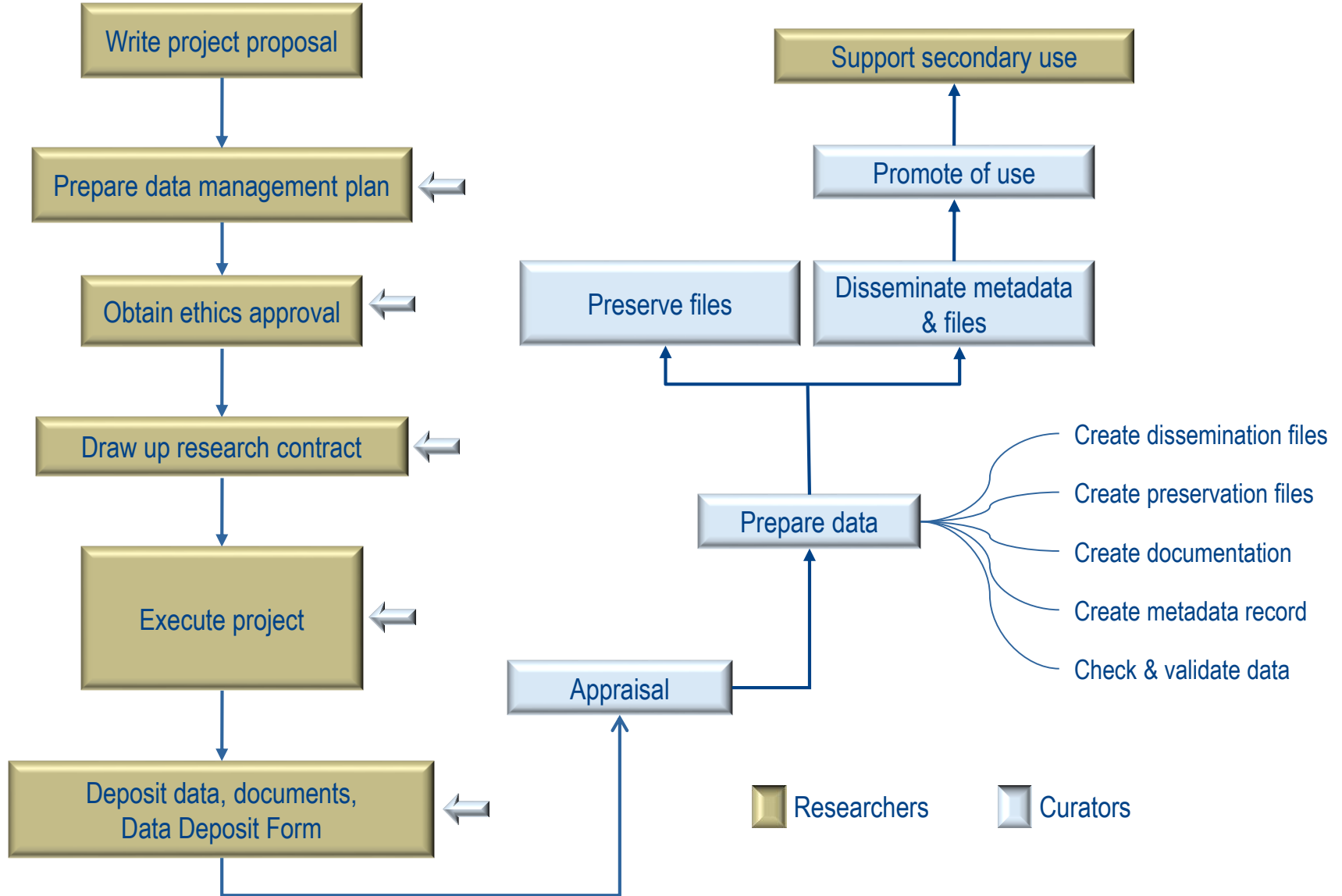
# What is the context of the curation process?

- **Outside-in**
  - What are the policies of the top journals in the domain?
  - What are legislative, regulatory requirements?
  - What is the research culture like?
  - What are the policies of funders of research in the domain?
- **Inside-out**
  - What resources, technology, capacity can be used?

# At the HSRC ...

- Policies and procedures that facilitate data deposit, preparing data and related documentation and data sharing
- Support for researchers in terms of data curation issues
- Training of researchers in data documentation and management
- A metadata and file repository
- An on-line dissemination interface linked to the HSRC's Web Portal for viewing, downloading or analysis
- Processes to monitor and audit curated data sets for performance information purposes
- Various data sets available for secondary use

# Data management process flow



# Data flow

Shared storage area

you are here: [browse](#) » [folders](#) » [data curation](#) » [sasas](#) » [sasas 2003](#)

Enter search criteria...

**About this folder**

- Display Details
- Folder transactions

---

**Actions on this folder**

Upload Document

Title	Created	Modified	Creator	Workflow State
Open Access	—	—	Lucia L. Lotter	
Registered Access				
Restricted Access				

Title	Created
SASAS2003 Consent_adolescent (62Kb)	2010-07-28 15:16
SASAS2003 Consent_adult (108Kb)	2010-07-28 15:17
SASAS2003 Letter_of_Introduction (50Kb)	2010-07-28 15:46

3 items, 25 per page

Delete Move Copy Archive Download All Checkout

File repository

Metadata repository

Data Deposit Form

1

**Data sets interface - Edit data set**

Data set ID: SASAS 2003 Q1 13/255

Control panel Data set details Funding / Authoring Scope Data collection

Title: South African Social Attitudes Survey (SASAS), 2003: Questionnaire 1 - Nine provinces in South Africa

Citation: Human Sciences Research Council. South African Social Attitudes Survey (SASAS) Round 1 2003 [Computer file]. SASAS Questionnaire 1. Pretoria South Africa: Human Sciences Research Council [producer] 2003, [distributor] 2007.

Data set

**South African Social Attitudes Survey (SASAS), 2003: Questionnaire 1 - Nine provinces in South Africa**

All data sets Data set details Documentation Data files Access conditions Contact

Data set metadata record

Data set ID : SASAS 2003 Q1

Title : South African Social Attitudes Survey (SASAS), 2003: Questionnaire 1 - Nine provinces in South Africa

Citation : Human Sciences Research Council. South African Social Attitudes Survey (SASAS) Round 1 2003 [Computer file]. SASAS Questionnaire 1. Pretoria South Africa: Human Sciences Research Council [producer] 2003, [distributor] 2007.

2  
Dissemination in Web Portal

Public open access  
Public registered access  
Limited interest group  
Project team  
Long term preservation

3  
Preservation



# .. Dissemination

## Data set

### South African HIV/AIDS, Behavioural Risks, Sero-status, and Mass Media Impact Survey (SABSSM), 2002: Adult and youth - Nine provinces in South Africa

All data sets	<b>Data set details</b> ▾	Documentation ▾	Data files	Outputs ▾	Access conditions	Contact
---------------	---------------------------	-----------------	------------	-----------	-------------------	---------

Data set metadata	Description	
	Funding / Authoring	
Data set ID :	Scope	2002 Adult-youth
	Data collection	
Title :	Subject information	South African HIV/AIDS, Behavioural Risks, Sero-status, and Mass Media Impact Survey (SABSSM), Adult and youth - Nine provinces in South Africa
	Metadata record	

**Citation :** Shisana, O. & Simbayi, L.C. *South African HIV/AIDS, Behavioural Risks, Sero-status, and Mass Media Impact Survey (SABSSM), 2002*. [Computer file]. SABSSM 2002 Adult-youth. Cape Town South Africa: Human Sciences Research Council, HIV/AIDS, STI and TB programme (HAST) [producer] 2002, [distributor] 2011

**Description :** The adult and youth data of the SABSSM 2002 study cover information from adults and youths 15 years and older on topics ranging from biographical information, media and communication, male circumcision, marital status and marriage practice, partner and partner characteristics, sexual behaviour and practices, voluntary counseling and testing (VCT), sexual orientation, interpersonal communication, practices around widowhood, knowledge and perceptions of HIV and AIDS, stigma, hospitalisation and health status.

**Abstract :** Background: This is the first in a series of national HIV household surveys conducted in South Africa. The survey was commissioned by the Nelson Mandela Children's Fund and the Nelson Mandela Foundation. The key aims were to determine the HIV prevalence in the general population, identify risk factors that






# .. Dissemination

## Data documentation

### Data documentation related to South African HIV/AIDS, Behavioural Risks, Sero-status, and Mass Media Impact Survey (SABSSM), 2002: Adult and youth - Nine provinces in South Africa

All data sets	Data set details ▾	<b>Documentation ▾</b>	Data files	Outputs ▾	Access conditions	Contact
<b>Data documentation</b>		Introductory information	 <i>Contextual documents</i>			
<b>Introductory information</b>		Data collection documents				
■ (2282) SABSSM2002_readme.txt 📄		Data input notes				
Data processing notes		Ethical research documents				
		Training manuals				

■ (2283) SABSSM2002\_Studyinformation.pdf 📄

This document provides information regarding bibliographic citation, copyright and disclaimer, acknowledgement in case of publications and the metadata record of the study.

■ (2281) SABSSM2002\_Userguide.pdf 📄

The user guide contains background information pertaining to the survey and It also has additional information and examples about the use of SPSS, Stata a



**Public open access:** Information is immediately available without the need to register, provide any additional information or obtain approval.



**Public registered access:** Information can be accessed after a user has registered and provided a reason for wanting access. No approval is necessary.



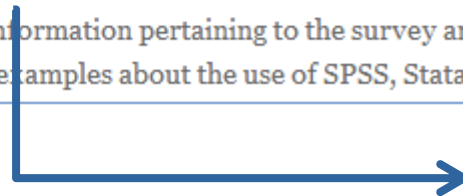
**Limited interest group:** Users are requested to register and provide a reason for wanting to access the information. Access is subject to approval from the owners, funders or depositors of the data. An email notification will be sent to confirm that access has been granted.



**Project team:** Specified project team members will have access to the information. If a user (who is part of the project team) was not entered as part of the project team the system will request the user to register and provide a reason for wanting to access the information. Access will be granted when approval is received.



*DC Metadata standard*



# .. Dissemination

## Data files

Data files related to South African HIV/AIDS, Behavioural Risks, Sero-status, and Mass Media Impact Survey (SABSSM), 2002: Adult and youth - Nine provinces in South Africa










*EUL*

## Data files

It is advisable to study the [introductory information](#) before using the data or related documents as it provides a systematic exposition of what the collection entails and how it should be used.

Note: Old versions of web browsers such as Internet Explorer 6 could cause problems when downloading files.

Download	Access	File	Description
ASCII FIXED FORMAT		<a href="#">SABSSM2002_ADULT_YOUTH.DAT</a>	Save file to disk and use as input file for the different programs.
SAS DATA SET		<a href="#">SABSSM2002_ADULT_YOUTH.SAS7BDAT</a>	Save file to disk. See <a href="#">userguide</a> on how to reference the data set and formats in a SAS program.
SAS FORMATS		<a href="#">SABSSM2002_ADULT_YOUTHFORMATS.SAS7B</a>	 <b>Public open access:</b> Information is immediately available without the need to register, provide any additional information or obtain approval.  <b>Public registered access:</b> Information can be accessed after a user has registered and provided a reason for wanting access. No approval is necessary.  <b>Limited interest group:</b> Users are requested to register and provide a reason for wanting to access the information. Access is subject to approval from the owners, funders or depositors of the data. An email notification will be sent to confirm that access has been granted.  <b>Project team:</b> Specified project team members will have access to the information. If a user (who is part of the project team) was not entered as part of the project team the system will request the user to register and provide a reason for wanting to access the information. Access will be granted when approval is received.

*Dissemination formats*

# .. Dissemination

## Research outputs


### Research outputs related to South African HIV/AIDS, Behavioural Risks, Sero-status, and Mass Media Impact Survey (SABSSM), 2002: Adult and youth - Nine provinces in South Africa

All data sets	Data set details ▾	Documentation ▾	Data files	<b>Outputs ▾</b>	Access conditions	Contact
---------------	--------------------	-----------------	------------	------------------	-------------------	---------

**Research outputs**

**Journal Article**

- Journal Article
- Monograph (Book)
- Review in Journal



*Outputs linked on data set level*

■ Connolly, C., Simbayi, L.C., Shanmugam, R. & Nqeketo, A. (2008) Male circumcision and its relationship to HIV infection in South Africa: results of a national survey in 2002. *South African Medical Journal*. 98(10):789-794.

the objective of this article is to investigate the nature of male circumcision and its relationship to HIV infection. Method used: analysis of a sub-sample of 3 025 men aged 15 years and older who participated in the first national population-based survey on HIV/AIDS in 2002. Chi-square tests and Wilcoxon rank sum tests were used to identify factors associated with circumcision and HIV status...

■ Shisana, O., Zungu-Dirwayi, N, Toefy, Y., Simbayi, L C, Malik, S. & Zuma, K. (2004) Marital status and risk of HIV infection in South Africa. *South African medical journal*. 94(7):537-543.

The relationship between marital status and HIV is complex. The risk depends on various demographic factors and sex behaviour practices. Increased prevention strategies that take socio-cultural context into account are needed for married people.

■ Shisana, O., Stoker, D., Simbayi, L.C., Orkin, M., Bezuidenhout, F., Jooste, S.E., Colvin, M. & Van Zyl, J. (2004) South African national household survey of HIV/AIDS prevalence, behavioural risks and mass media impact: detailed methodology and response rate results. *South African medical journal*. 94(4):283-287.

■ Simbayi, L.C., Chauveau, J. & Shisana, O. (2004) Behavioural responses of South African youth to the HIV/AIDS epidemic: a nationwide survey. *AIDS care*. 16(5):605-618.

#### Monograph (Book)

■ Brookes, H., Shisana, O. & Richter, L. (2004) *The national household HIV prevalence and risk survey of South African Children*. Cape Town: HSRC Publishers.

# Check and validate data

- Validation - error detection
- Cleaning - correcting errors
  - Done by researchers
  - Document issues
- Check anonymisation

# Create data documentation

- Create catalogue record
- Create a data listing for the study
- Enhance internal metadata (variable and value labels)
- Develop contextual information
- Catalogue related documentation
- Link research outputs to data

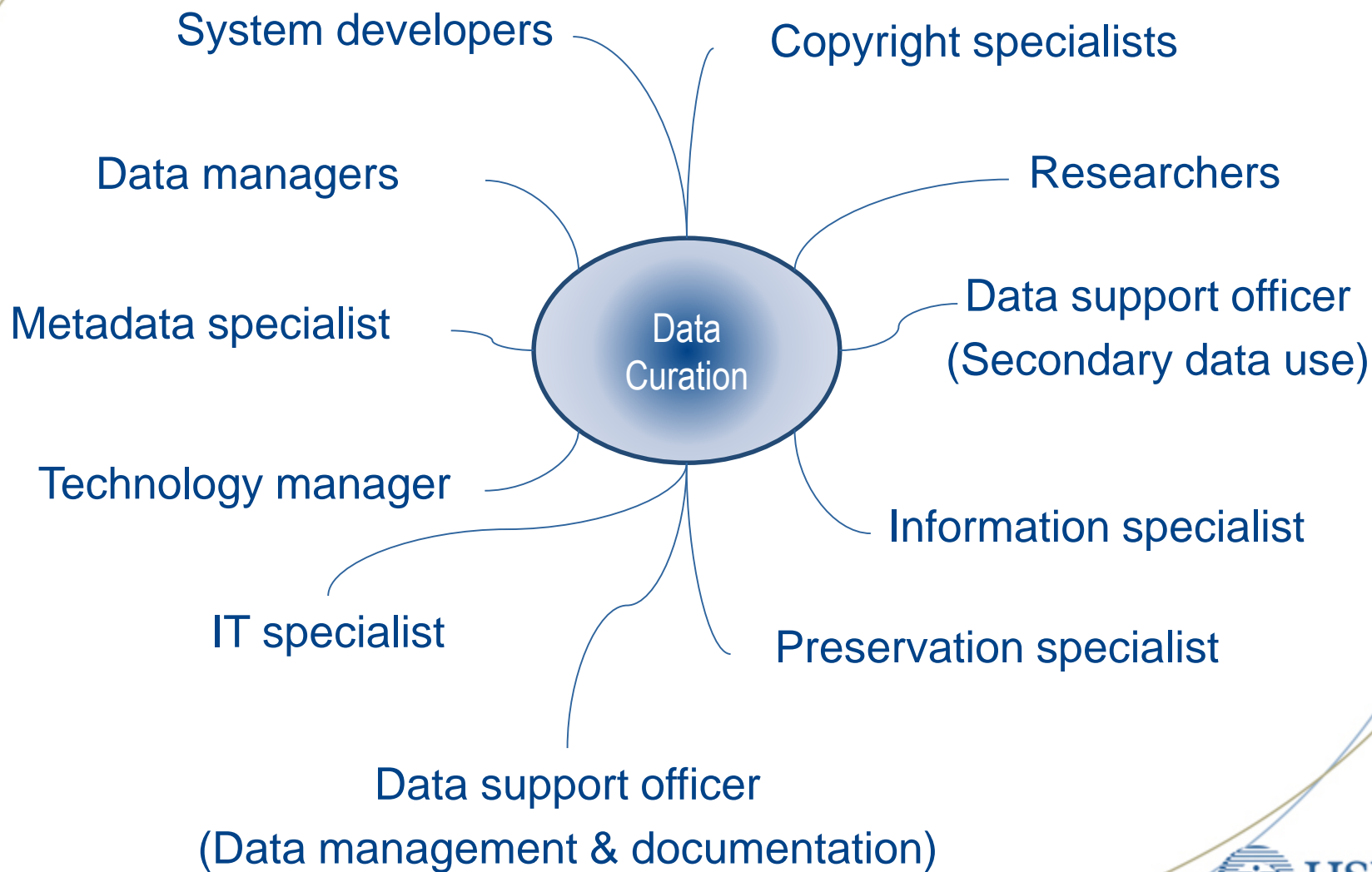
# Create preservation and dissemination files

- Format related challenges
  - Software obsolescence
  - Format conversion
- Dissemination formats
- Preservation formats

# Preservation

- Preservation planning
  - Preservation metadata
  - Monitoring file formats and media
- Preservation strategies
  - Storage and storage management (multiple copies, multi-media)
  - Back-up, disaster recovery
  - Security

# Many roles, Multiple skills





# Essential activities

- Engage with data producers - *the research staff*
- Develop facilitating workflows - *make it easy for everyone*
- Implement a suitable technology platform - *stable, reliable, scalable*
- Develop data curation policies - *standard management tools*
- Create and pilot service models - *look for 'tame' users*
- Do change management - *advocate, advocate, advocate!*

Walters, T.O. (2009). Data Curation Program Development in U.S. Universities:  
The Georgia Institute of Technology Example. The International Journal of Digital Curation 3(4):83-92.  
[www.ijdc.net/index.php/ijdc/article/view/136/153](http://www.ijdc.net/index.php/ijdc/article/view/136/153)

# Data dissemination - examples

The Collaborative HIV/AIDS and Adolescent Mental Health Project (CHAMP)

<http://www.hsrc.ac.za/en/projects/view/SAIAAA>

Audit of technological initiatives for rural development

<http://www.hsrc.ac.za/en/projects/view/MJAFAA>

Baseline patient satisfaction survey in 275 clinics located in three health districts

<http://www.hsrc.ac.za/en/projects/view/PGAZMA>

HIV/AIDS educators study

<http://www.hsrc.ac.za/en/projects/view/PCMABA>

The HIV and Alcohol Prevention in Schools (HAPS)

<http://www.hsrc.ac.za/en/projects/view/SBELBA>

Mentor mothers programme

<http://www.hsrc.ac.za/en/projects/view/PJAQNA>

Literacy and numeracy research

<http://www.hsrc.ac.za/en/projects/view/LFATBA>

The Status of Youth Report 2003

<http://www.hsrc.ac.za/en/projects/view/SCILAA>

Non-conforming gender identities

<http://www.hsrc.ac.za/en/projects/view/SLAFAA>

Socio-Economic Impact of the Global Financial Crisis on South Africa

<http://www.hsrc.ac.za/en/projects/view/MIAHAA>

South African HIV/AIDS Behavioural Risks, Sero-Status, and Mass Media Impact Survey

<http://www.hsrc.ac.za/en/projects/view/PFAJLA>

South African Social Attitudes Survey (SASAS)

<http://www.hsrc.ac.za/en/projects/view/TAAMAA>

Trends in International Mathematics and Science Study 2003

<http://www.hsrc.ac.za/en/projects/view/LAAQBA>

Raising visibility : WSW, LBT communities experiences in HIV and AIDS

<http://www.hsrc.ac.za/en/projects/view/SIAUAA>

Young Fathers: Experiences of fatherhood in the context of poverty in SA –  
A phenomenological and ecological investigation

<http://www.hsrc.ac.za/en/projects/view/SGASPA>



Thank you

**Building the bridge between  
research, policy and action**

Lucia Lötter  
lloetter@hsrc.ac.za  
www.hsrc.ac.za